УДК 004.8:004.932

***М.В. Давидов, Ю.В. Нікольський***
Національний університет «Львівська політехніка», м. Львів, Україна
maxd128@mail333.com, y_nikol@yahoo.com

# Методи та засоби опрацювання зображень реального часу для ідентифікації елементів жестової мови

Об'єктом досліджень є жестова мова, якою спілкуються люди з пониженим слухом. Мета проведених досліджень – створення комп'ютерної системи для спілкування з такою категорією людей. Розглянуто два різновиди мови – дактиль, жести якого зображають букви алфавіту та цифри, та повна жестова мова. Проведено порівняння різних способів виділення ознак жесту. Запропоновано підхід до виділення ознак жесту введенням еталона. Досліджено ефективність побудованих алгоритмів для розпізнавання жесту в реальному часі. Пропонований підхід реалізовано у вигляді прототипу тренажера для навчання мови жестів.

## Вступ

Люди з пониженим слухом спілкуються між собою спеціальною жестовою мовою, яка складається з наборів жестів. Спілкування з такими людьми вимагає залучення сурдоперекладачів, яких є надзвичайно мало. Підтвердженням актуальності цього є доручення, яке надане президентом України В. Ющенком міністру освіти і науки про початок підготовки сурдоперекладачів у вищих навчальних закладах України з 1 вересня 2008 року (http://unian.net/ukr/news/news-224756.html). Спілкування двох осіб, одна з яких може володіти жестовою мовою, а друга – ні, обмежує можливості до спілкування людей з пониженим слухом. Важливою задачею є вивчення жестової мови дорослими людьми, які втратили слух. Особливої гостроти набуває проблема такого спілкування у разі участі цих людей у дискусіях та виступах на конференціях, навчання в спеціалізованих школах. Тому значної актуальності набуває розроблення нових засобів спілкування на основі сучасних інформаційних технологій. Ми надаємо огляд результатів досліджень [1-4], які виконані з метою створення комп'ютерно-програмних систем для допомоги у спілкуванні українською жестовою мовою. Для цього виконано аналіз систем візуальних сигналів та розроблено програмні засоби перетворення цих сигналів у повідомлення. Особлива увага приділялась можливості використання такої системи у реальному часі. Розроблення системи виконувалось із врахуванням специфіки української жестової мови.

## Мета роботи

Проведене комплексне дослідження аспектів виділення ознак елементів жестової мови для побудови комп'ютерно-інформаційної системи ідентифікації жестів. Для досліджень використано підручник української жестової мови у форматі відеофільму, з допомогою якого навчають жестової мови у спеціалізованих школах для дітей з вадами слуху. Для ідентифікації жестів проаналізовано відеозображення та створено прототип програмно-апаратного комплексу, з допомогою якого можна

навчати жестової мови людей, що втратили слух або вивчають мову жестів для спілкування. Повний комплекс досліджень передбачає розроблення математичних моделей жесту, алгоритмів та програмних рішень, розпізнавання жесту на зображенні та його ідентифікацію. Проведені дослідження дозволили розв'язати такі задачі:

– проаналізувати особливості кодування відеозображень елементів жесту;
– виконати розпізнавання кінців пальців руки для ідентифікації елементів дактиля, з допомогою якої зображають букви алфавіту для транслітерації прізвищ, термінів, скорочень тощо;
– побудувати алгоритми розпізнавання форми кисті руки на відеозображенні у динаміці та у реальному часі.

Вказані задачі сформульовано у термінах задач прийняття рішень. Для їх вирішення застосовано методи машинного навчання. Математичне моделювання динамічних об'єктів також пов'язане із побудовою систем, які навчаються, та відповідних алгоритмів машинного навчання. Математичні моделі передбачали розпізнавання динамічних об'єктів у реальному часі. Вказаними об'єктами є відеозображення, які проходять оброблення з метою виявлення та ідентифікації елементів мови жестів. Дослідження жесту з метою побудови прототипу комп'ютерної системи ідентифікації дозволило:

– ввести поняття елемента жесту, основної конфігурації та її еталона;
– розробити спосіб створення еталонів;
– автоматично розпізнавати кінці пальців руки на кадрах відеозображення;
– розробити метод знаходження форми кисті, близької до еталона, а на цій основі побудувати алгоритм розпізнавання жесту за формою кисті руки;
– розробити алгоритм знаходження та ідентифікації форми кисті руки на відеокадрі;
– визначити підходи до вирішення задачі ідентифікації елементів жестової мови.

# Постановка задачі досліджень

Словник української жестової мови налічує приблизно дві тисячі жестів. Кожний жест означає букву, слово або словосполучення та складається з певних елементів, які задані фіксованими положеннями частин тіла. Жести можна розглядати нерухомими або в русі, а також такими, що виконують пальцями рук, усією долонею, усією рукою, обома руками, мімікою обличчя, рухами частин тіла та артикуляцією губами. Ми розглядали елементи жестів. У подальшому до них буде додано рухи рук, міміку обличчя, артикуляцію слів губами, рухи частин тіла тощо.

Для вирішення проблеми розпізнавання жесту в науковій літературі описане застосування методів головних компонент, нечітких нейронних мереж, псевдодвовимірних марківських моделей, методів на основі порівняння зображень або порівняння зі спотвореннями, методу моделі тощо. Метод головних компонент [5] використано для розпізнавання форми кисті руки. Автори стверджують, що досягли 89 % точності розпізнавання для 100 різних жестів, виконаних в динаміці. Такий метод виявився чутливим до локальних спотворень зображень: навіть невеликий шум на зображенні або його деформація суттєво впливають на результат розпізнавання. Висока швидкість обчислення за методом головних компонент дозволяє відносити зображення до певної групи, а вже точну ідентифікацію форми кисті здійснювати іншим методом. Для методу нечітких нейронних мереж [6] характерне

розпізнавання невеликої кількості класів зображень. За твердженням авторів методу, ними досягнуто 92 % правильних розпізнань форми кисті руки для статичного зображення жестів. Для класифікації зображень, які розтягують або стискають, добре пристосований метод псевдодвовимірних марківських моделей [7]. Мала швидкодія цього методу ускладнює його використання для класифікації великої кількості жестів. Методи порівняння зображень [8] використовують різні способи оцінювання їх відмінності. Найпоширенішими є міри відмінності для попіксельного порівняння зображень. Результати застосування методу порівняння зображень зі спотвореннями [9] не залежать від невеликих зміщень кожного піксела. Недоліком цього методу є те, що подібними в процесі порівняння тут вважають зображення, на яких змінюється послідовність кольорів пікселів. Висока швидкодія методу дозволяє використати його для опрацювання відеозображень реального часу. У роботі [10] розпізнавання руки на зображенні здійснено за методом моделі, який вимагає попереднього відділення пікселів руки від пікселів тла. Це дозволяє використовувати його для розпізнавання лише найпростіших форм кисті руки. Перевага методу полягає в тому, що результати його застосування не залежать від зміни форми проекції кисті руки під час її повороту. Аналіз запропонованих методів показав, що жоден з них не дає достатньої якості розпізнавання розмитих рухомих зображень пальців та кисті руки. Також усі вони мають низьку швидкодію. Тому проведено дослідження, які присвячені вирішенню задач, пов'язаних із підвищенням якості та швидкодії розпізнавання жесту.

## Опрацювання зображення жесту та його ідентифікація

Результати проведених досліджень складаються з алгоритмів виділення ознак жестів, які виконують однією рукою з метою побудови алгоритмів встановлення відповідності жестової мови буквам, словам та реченням. Проведені дослідження реалізовано у вигляді спеціалізованого програмного забезпечення, яке є основою створення комп'ютерно-програмної системи для спілкування з людьми з пониженим слухом. Задача ідентифікації жесту за його елементами вирішувалась як дві групи досліджень [1-4]. Перша група полягає у побудові математичної моделі, яку використано для розпізнавання кінців пальців на зображенні та виділення їх серед інших елементів зображення. В результаті цих досліджень здійснено розпізнавання на зображенні пальців однієї руки та конфігурацій, ними утворюваних, що дозволяє класифікувати такі конфігурації та ідентифікувати елементи дактиля. Приклади розпізнаних кінців пальців показано на рис. 1, який виконаний на спеціально опрацьованих відеозображеннях реального часу. Це дозволяє вирішувати задачу класифікації комбінацій пальців та ставити їм у відповідність букви дактиля. Друга група досліджень пов'язана із ідентифікацією жестів підмножини жестової мови, яка складалась з 12 жестів. Тут здійснено розпізнавання та ідентифікацію кисті руки, якою виконують жест. Жест на відеозображенні розглядається як послідовність кадрів, на яких відбувається зміна форми кисті руки. Класифікації жесту передує пошук кисті руки на відеокадрах із зображенням людини, що жестикулює. Після цього встановлюється відповідність між формою кисті руки та певним попередньо створеним еталоном.



Рисунок 1 – Приклади кадрів із розпізнаними кінцями пальців

Математичною моделлю задачі ідентифікації ознак жесту шляхом розпізнавання кінців пальців та форми кисті руки є система прийняття рішень у такому формулюванні [11]: $B = (Z, A \bigcup \{d\})$, де $Z$ – множина об'єктів, $A = \{ a_1, a_2, ..., a_m \}$ – множина атрибутів, які встановлюють відповідність вигляду $a_i : Z \to V_a$ , $i = 1,2,...,m$ , де $V_a$ – множина значень атрибута $a_i$, $d$ – атрибут прийняття рішень. Атрибути множини $A$ називають умовними або умовами, а $d$ – рішенням. Кортеж $z = \{ z_1, z_2, ..., z_m \}$, $z \in Z$ , складається зі значень відповідних атрибутів об'єкта $z$ . Спочатку наведемо результати розв'язання задач, які стосуються першої групи досліджень. Відповідний алгоритм детально описано у роботі [1]. Цей та інші алгоритми тут не наводимо внаслідок обмежень на розмір статті. Традиційний підхід до розв'язання задач розпізнавання полягає у створенні шаблонів, які потім шукають на зображенні. Оскільки положення пальців, їх розмір та орієнтація можуть змінюватись на відеозображенні реального часу в широких діапазонах, то традиційний підхід вимагає великої кількості шаблонів. У задачі розпізнавання кінців пальців об'єктом системи прийняття рішень є точка зображення, його атрибутами – значення яскравості пікселів з її околу, атрибут прийняття рішень набуває значення одиниці, якщо точка належить кінцю пальця, і нулю – якщо не належить. Для оптимізації обчислень обираються точки, які є центрами пікселів. Для ідентифікації елементів зображення використано нейромережу, побудовану за схемою багатошарового персептрона, навчання якої виконано за модифікованим методом зворотного поширення похибки [1], [12]. Експериментально підібрано параметри цієї мережі – один прихований шар з п'ятьма нейронами та непарною активаційною функцією. Такі параметри мережі дозволили отримати до 90 % правильних розпізнавань кінців пальців для вибраних умов відеозйомки та способу підготовки зображення. Для уточнення синаптичних зв'язків мережі алгоритм навчання використовує відмінність навчального прикладу від побудованого еталона. Задачу класифікації вирішено у два етапи: на першому з них на навчальних прикладах навчено мережу, а на другому – таку мережу використано для класифікації нових прикладів. До початку навчання створено навчальні приклади та відповідні їм еталони, які мають дещо інший зміст, ніж шаблони у традиційних задачах розпізнавання. Приклади та еталони створені з кадрів відеозображення опрацюванням відеокадрів, що описано в [1].

Опрацювання кадрів здійснено з метою зменшення обсягу інформації про зображення. Навчальні приклади сформовані з пікселів зображення, які належать кінцям пальців. Для цього на частині кадрів були відзначені кінці пальців, а їх

розпізнавання розглядалось як поділ пікселів зображення на дві групи: одна з яких містила піксели, які належать кінцям пальців, а друга – такі, які не належать. Нейронна мережа обчислює функцію, значеннями аргументів якої є компоненти прикладів, а значенням функції – ступінь належності піксела кінцю пальця. Навчена нейромережа розпізнала кінці пальців на новому зображенні так, як це показано на рис. 1. Для цього кожному пікселу зображення поставлений у відповідність приклад, який подано на вхід мережі, а мережа обчислила ступінь належності цього піксела кінцю пальця. Друга група досліджень стосувалась розпізнавання форми кисті руки, якою виконують жест. Розпізнавання виконувала програма за спеціальним алгоритмом, ідея якого полягає у створенні множини основних форм кисті та відповідної множини еталонів. Для прикладу на кадрах з рис. 2а – 2б показано виконання жесту «прибуток». Під час виконання цього жесту зображення кисті руки приймає 5 різних форм. Тому виконання жесту можна представляти як послідовність таких форм, які назвемо основними. Підхід, який базується на понятті основних форм, дозволяє створити усі основні форми жестової мови. Тоді кожний жест на кадрах відеозображень можна зображати послідовністю форм, подібних основним. Основні форми створює людина, вибираючи їх на зображеннях або демонструючи перед камерою. Створення основних форм можна виконати також кластеризацією об’єктів на кадрах відеозображень. Задача розпізнавання жесту ускладнена необхідністю враховувати індивідуальні риси співрозмовника: колір шкіри, форму кисті руки, швидкість рухів тощо – та умови зйомки: освітлення, колір та малюнок тла. Суттєве ускладнення алгоритмів обробки пов’язане із необхідністю опрацьовувати завади, викликані недоліками відеокамери: розмиттям зображення, неадекватністю відтворення кольорів, засвітленням або затемненням частин зображення, похибками дискретизації тощо.

Побудовано алгоритм знаходження міри подібності зображень, які вважаємо отриманими одне з другого в результаті невеликих деформацій та зміни кольорів окремих елементів. Цей алгоритм є модифікацією методу, який використовує псевдодвовимірні приховані марківські моделі [13]. Наш алгоритм дозволив досягнути прийнятних результатів класифікації за допомогою одного прикладу кожного класу. Розроблений метод використовує навчання з підкріпленням [14] та знаходить міру подібності елементів зображень розв’язанням задачі динамічного програмування. На жаль, розміри цієї статті не дозволяють викласти повністю алгоритм розв’язання. Ідентифікацію жесту руки за формою кисті розглядали як задачу класифікації. Для її розв’язання введено ознаки класів об’єктів, а результат розпізнавання полягає у наданні досліджуваним об’єктам цих ознак. Ознаками класів у випадку розпізнавання жестів є змістовна назва жесту, а віднесення до класу виконується обчисленням міри подібності зображень та оцінюванням близькості об’єктів за значеннями цієї міри. Пропонований метод використовує еталони, поставлені у відповідність кожній основній формі. Еталоном називаємо прямокутну область, яка складається з пікселів із зображенням кисті руки в основній формі та прозорих пікселів тла. Усі еталони мають однакові розміри. Пошук за еталоном полягає у знаходженні на зображенні прямокутної області, яка серед інших прямокутних областей такого ж розміру найподібніша одному з еталонів за найбільшим значенням міри подібності.




Рисунок 2а – Кадри відеофільму із зображенням жесту «прибуток»

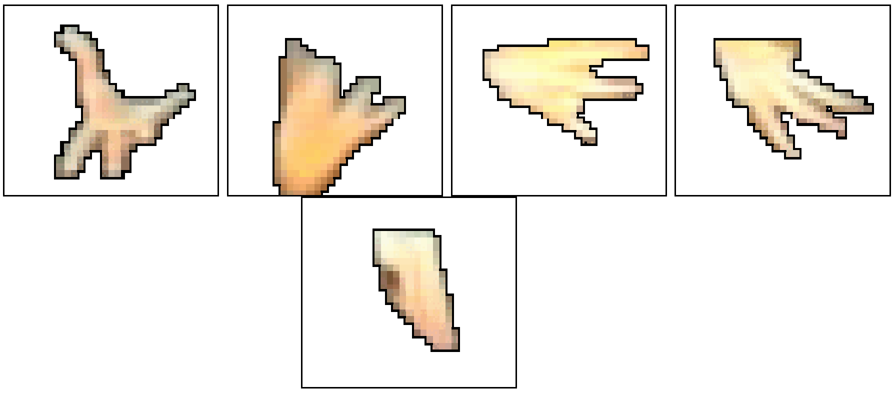


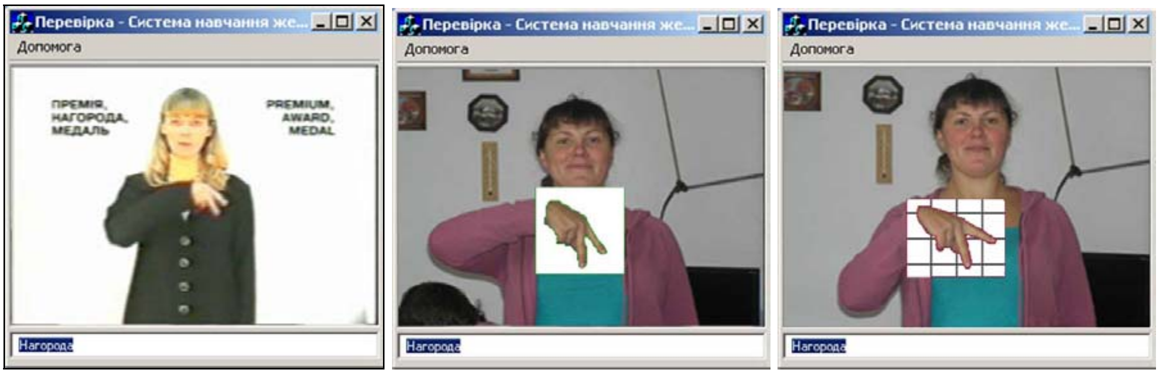
Рисунок 2б – Еталони, за якими відбувається ідентифікація жесту «прибуток»

Подібними вважаємо зображення, отримані одне з одного невеликою трансформацією, якою є невелика зміна кольору, переміщення, локальні стиск або розтягнення його частин. Величину трансформації задаємо значеннями спеціальних параметрів. Врахування відносно невеликих деформацій та змін кольорів зображення кисті руки при виконанні жесту підвищило якість розпізнавання та скоротило на 62,5 % кількість помилок із одночасним збільшенням на 67 % часу обчислень на відміну від попіксельного порівняння із шаблоном.

Результати проведених досліджень покладені в основу створення тренажера для навчання жестової мови. Пропонований тренажер містить словник жестової мови у вигляді множини відеокадрів (рис. 3а) та засоби керування переглядом відеокадрів. Також тренажер дозволяє контролювати процес навчання мови жестів. Спеціальні засоби дозволяють учневі бачити на екрані комп'ютера своє зображення, що надходить з відеокамери. Це дозволяє синхронізувати виконання жесту учнем із демонстрацією цього жесту зі словника. Функції, які закладено у тренажер, дозволяють реалізувати розроблені алгоритми розпізнавання жесту, встановлювати правильність виконання жесту та сигналізувати про це учневі. На рис. 3б відмічено правильне виконання жесту, а на рис. 3в – неправильне. Важливою особливістю роботи тренажера є можливість ідентифікувати жест у реальному часі в процесі його виконання.

# Висновки

Побудовано та досліджено алгоритми виділення ознак жесту, виконано алгоритмічну та програмну реалізацію методів його ідентифікації. Це використано для розпізнавання кінців пальців у разі спілкування дактилем та ідентифікації кисті руки за еталоном. Введення еталонів дозволило розглядати жест як послідовність невеликої кількості основних форм.




а б в

Рисунок 3 – Вікна тренажера для вивчання жесту «нагорода»: а) кадр із зображенням жесту «нагорода» із відеословника жестової мови; б) правильне виконання жесту учнем (рука виділена білим квадратом); в) неправильне виконання жесту учнем (рука виділена штрихованим квадратом)

Такий підхід дозволив досліджувати жест як різновид невеликої кількості базових елементів, що зменшило кількість форм, які аналізувались під час ідентифікації елементів жесту. Реалізація алгоритмів із введеним поняттям основної форми дозволила зображати кожну форму одним або кількома еталонами, а кожний елемент жестової мови – скінченною послідовністю еталонів. Це надало можливість шукати зображення кисті руки, найподібніше до еталона. Проведені експерименти показали ефективність запропонованих підходів та дозволили оцінити його чутливість у разі врахування деформацій об’єктів та зміни кольорів на зображеннях. Запропонований метод розпізнавання вимагає лише одного прикладу для опису елемента жесту.

# Література

1. Давидов М.В., Нікольський Ю.В. Класифікація елементів відеозображень реального часу з допомогою нейромережі // Вісник Національного університету «Львівська політехніка»: Інформаційні системи та мережі. – 2005. – № 549. – С. 82-92.
2. Давидов М.В., Нікольський Ю.В. Нейромережний класифікатор елементів відеозображень реального часу // Вісник Національного університету «Львівська політехніка»: Комп’ютерні системи проектування. Теорія і практика. – 2006. – № 564. – С. 18-25.
3. Davydov M., Nikolski I., Pasichnyk O. System of finger movement identification for sign language recognition // Abstracts of First Central European Student Conference in Linguistics, 29 – 31 May 2006. – Budapest (Hungary). – P. 23-25.
4. Давидов М.В., Нікольський Ю.В., Пасічник В.В. Математичне моделювання та програмна реалізація елементів тренажера для навчання жестовій мові людей, що втратили слух // Труды Седьмой Международной конференции «Интеллектуальный анализ информации (ИАИ-2007)». – Киев, 15 – 18 мая 2007 г. – С. 56-66.
5. Atid Shamaie, Alistair Sutherland. Accurate Recognition of Large Number of Hand Gesture. IranMVIP2003.pdf (2003).

6. Nguyen Dang Binh, Toshiaki Ejima. Hand Gesture Recognition Using Fuzzy Neural Network // GVIP 05 Conference, 19 – 21 December 2005. – CICC, Cairo (Egypt).
7. Nguyen Dang Binh, Enokida Shuichi, Toshiaki Ejima. Real-Time Hand Tracking and Gesture Recognition System // GVIP 05 Conference, 19 – 21 December 2005. – CICC, Cairo (Egypt).
8. Sánchez-Nielsen E., Antón-Canalís L., Hernández-Tejera M. Hand Gesture Recognition for Human-Machine Interaction // Journal of WSCG. – Vol. 12, № 1-3. – ISSN 1213-6972. – WSCG'2004, February 2-6, 2003. – Plzen (Czech Republic).
9. Dreuw P., Keysers D., Deselaers T., Ney N. Gesture Recognition Using Image Comparison Methods // GW05. – 2005. – P. 124-128.
10. Bretzner L., Laptev I., Lindeberg T. Hand Gesture Recognition Using Multi-Scale Color Features, Hierarchical Models and Particle Filtering // Proceedings of the Fifth IEEE International Conference on Automatic Face and Gesture Recognition (FGR02). – 2002.
11. Komorowski J., Polkowski L., Skowron A. Rough Sets: A Tutorial // Eds. S.K. Pal and A. Skowron, Rough Fuzzy Hybridization: A New Trend in Decision-Making. − Springer-Verlag. − 1998. – P. 3-98.
12. Горбань А.Н. Обобщенная аппроксимационная теорема и вычислительные возможности нейронных сетей // Сибирский журнал вычислительной математики. – 1998. – Т. 1, № 1. – С. 12-24.
13. Rabiner L.R. A tutorial on Hidden Markov Models and selected applications in speech recognition // Proceedings of the IEEE. – 1989. – Vol. 77(2). – P. 257-285.
14. Рассел С., Норвиг П. Искусственный интеллект. – М.: Издательский дом «Вильямс», 2006.

***М.В. Давыдов, Ю.В. Никольский***

**Методы и средства обработки изображений реального времени для идентификации элементов жестового языка**

Объектом исследований является жестовый язык, на котором общаются люди со сниженным слухом. Цель проведенных исследований – создание компьютерной системы для общения с такой категорией людей. Рассмотрены две разновидности языка – дактиль, жесты которого изображают буквы алфавита и цифры, и полный жестовый язык. Проведено сравнение разных способов выделения признаков жеста. Предложен подход к выделению признаков жеста введением эталона. Исследована эффективность построенных алгоритмов для распознавания жеста в реальном времени. Предложенный подход реализован в виде прототипа тренажера для изучения языка жестов.

*Стаття надійшла до редакції 10.12.2007.*